Supplements

Table 1: Top 20 Features Selected

Selected Features
CLDN7|Claudin-7
CCNB1|Cyclin_B1
ESR1|ER-alpha
GATA3|GATA3
IGFBP2|IGFBP2
PRKCA |PKC-alpha
FASN|FASN
GAPDH|GAPDH
MYH11|MYH11
NDRG1|NDRG1_pT346
TFRC|TFRC
TUBA1B|Acetyl-a-Tubulin-Lys40
CDKN2A|P16INK4A;CDKN2A|p16_INK4a
ADAM6_x
CEACAM5_x
COL1A1_x
GFAP_x
KLK3_x
KRT14_x
KRT5_x
KRT6A_x
SFTPB_x

Table 2: Grid Search Parameter

Model	Paramaters	Tested Values
SVM	C	0.0001, 0.001, 0.01, 0.1, 1.0, 10.0, 100.0, 1000.0
RF	Maximum Number of Trees	10,20,30,40,50,60,70
	Number ofParameters	5,10,20
SDAE	Epochs	3,10,50,100,500,1000
	Learning Rate	1.0,0.1,0.01,0.001,0.0001,0.00001
	Batch Size	10,20,40
	Hidden Layer Sizes	100, 500, 1000
	Noise	0.0,0.01,0.2,0.4,0.5
Lasso	alpha	0.0001 , 0.0001269 , 0.00016103, 0.00020434, 0.00025929, 0.00032903, 0.00041753, 0.00052983, 0.00067234, 0.00085317, 0.00108264, 0.00137382, 0.00174333, 0.00221222, 0.00280722, 0.00356225, 0.00452035, 0.00573615, 0.00727895, 0.00923671, 0.01172102, 0.01487352, 0.01887392, 0.02395027, 0.03039195, 0.0385662 , 0.04893901, 0.06210169, 0.07880463, 0.1

Figure1:Process Flow